New developments in modeling hogs production and growth at a finer temporal resolution

Luca Sartore[1,2], Yijun Wei[1,2], Emilola Abayomi[2], Seth Riggins[2], Gavin Corral[2], Nell Sedransk[1,2]

[1]National Institute of Statistical Science (NISS)
[2]United States Department of Agriculture
National Agricultural Statistics Service (USDA NASS)

lsartore@niss.org

QPRC 2019
Applications of data science in diverse fields
June 12, 2019

Disclaimer

The findings and conclusions in this presentation are those of the author(s) and should not be construed to represent any official USDA of U.S. Government determination or policy.

Presentation outline

1. Swine survey at NASS
2. Swine industry characterization
3. Inventory equations
4. Case study
5. Conclusion

Swine survey at NASS

NASS conducts quarterly surveys to evaluate status of US swine industry

Reported quantities over previous three months:

- Monthly **pig crop**
- Monthly **sows farrowed**

Reported inventories at the first day of the surveyed quarter:

- **Breeding herd**
- **Market hogs** distinct in weight groups:
  1. Less than 50 lbs
  2. 50-119 lbs
  3. 120-179 lbs
  4. More than 180 lbs

Pros and cons of current modeling approaches

Constrained Kalman filter model

- Reliable during periods of equilibrium
- Too rigid during periods of shocks

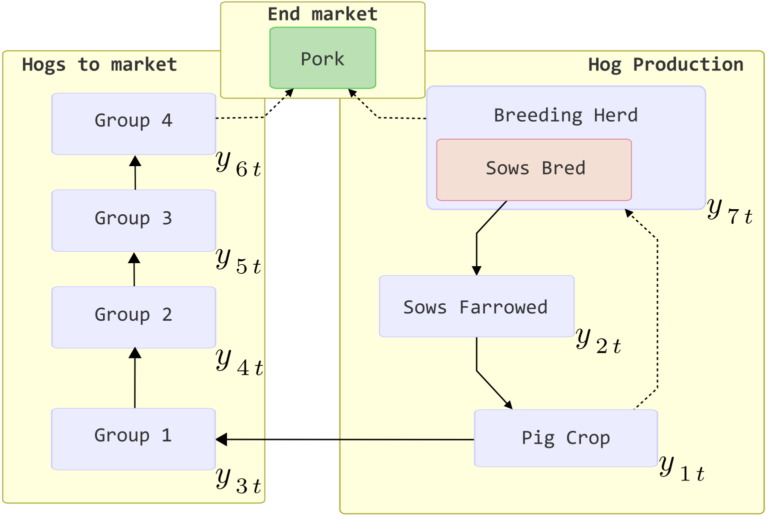
Sequential generalized linear models

- Very flexible
- Ignoring biological constraints

Objectives of the new model

1. Satisfy constraints without forcing them
2. Produce stable estimates during equilibrium periods
3. Flexibile formulation to quickly adapt estimates during shocks
4. Predict national inventories at a finer temporal resolution

Swine population dynamics



Breeding dynamics

$$\begin{cases} z_{1,t} = \rho_t \, z_{2,t} \\ z_{2,t} = \varphi \, y_{7,t-2} + \varepsilon_{2,t} \end{cases}$$

$z_{1,t}$ monthly pig-crop at time t

$z_{2,t}$ monthly sows farrowed at time t

$y_{7,t-2}$ size of breeding herd (inventory) at time $t-2$

ρ_t piglets per sows at time t

φ breeding/farrowing rate

$\varepsilon_{2,t}$ error in modeling sows farrowed

$$y_{7,t} = \varphi^{-1} \left(z_{2,t+2} + \varepsilon_{2,t+2} \right)$$

Time series models

Account for trend, seasonality, and auto-correlations

Dynamics of logarithms of pig crop and sows farrowed are modeled by SARIMA model (Box et al., 2015)

The model choice for $\log(z_{1,t})$ and $\log(z_{2,t})$ is

$$\text{SARIMA}(2, 1, 2) \times (2, 1, 2)_{12}$$

Model fitting via LASSO (Tibshirani, 1996) for automatic selection of a sub-model

Inventory equations

$$
\begin{cases}
y_{3,t} = \zeta_{t-1}\, z_{1,t-1} + \zeta_{t-2}\, z_{1,t-2} + \zeta_{t-3}\, \alpha_1\, z_{1,t-3} + \varepsilon_{3,t}, \\
y_{4,t} = \zeta_{t-3}\, (1-\alpha_1)\, z_{1,t-3} + \zeta_{t-4}\, z_{1,t-4} + \zeta_{t-5}\, \alpha_2\, z_{1,t-5} + \varepsilon_{4,t}, \\
y_{5,t} = \zeta_{t-5}\, (1-\alpha_2)\, z_{1,t-5} + \zeta_{t-6}\, \alpha_3\, z_{1,t-6} + \varepsilon_{5,t}, \\
y_{6,t} = \zeta_{t-6}\, (1-\alpha_3)\, z_{1,t-6} + \zeta_{t-7}\, \alpha_4\, z_{1,t-7} + \varepsilon_{6,t},
\end{cases}
$$

$z_{1,t}$ monthly pig-crop at time t

$y_{2+i,t}$ market hogs , for weight group $i = 1, \ldots, 4$

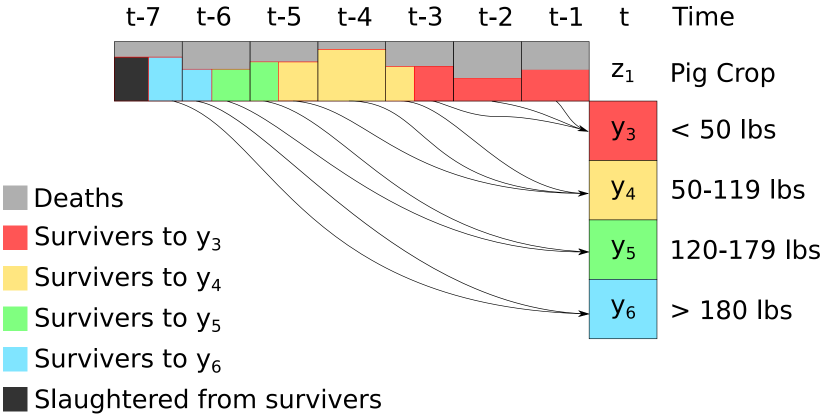
$\varepsilon_{2+i,t}$ errors in modeling market hogs

α_i projection parameters, for $i = 1, \ldots, 4$

ζ_t survival rates at time t

Inventory equations (cartoon interpretation)

Death rates and allocation percentages are not realistic!



Case study

Historical estimates from 2008 are combined with survey data up to 2017

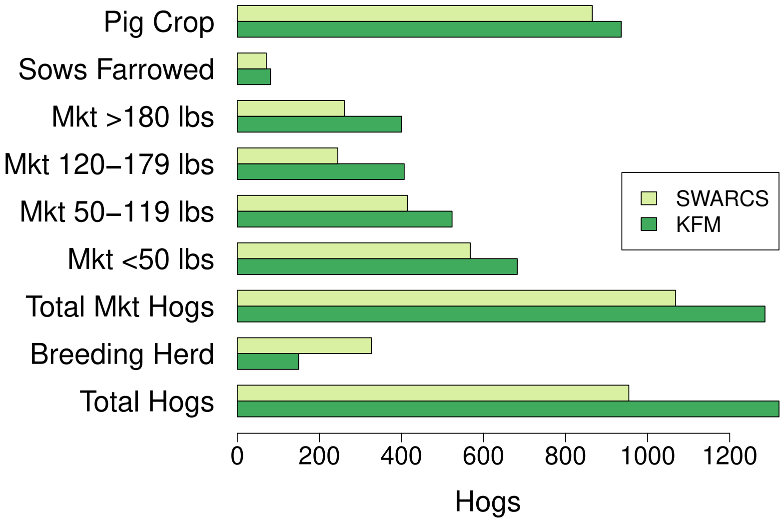
The proposed model (SWARCS) and the Kalman filter (KFM) are both used to estimate inventory numbers between 2013 and 2017

The results from the two models are compared against

- Final estimates[1] (NASS USDA, 2019)
- Root mean squared error (RMSE) as selection criterion

[1] https://downloads.usda.library.cornell.edu/usda-esmis/files/jd472w45t/h128nn160/m613n493n/hgpgsb19.pdf

Model comparisons via RMSE



Conclusion

For the new model

1. Biological dynamics are satisfied without forcing constraints
2. Stable estimates are produced by a flexible formulation of survival rates
3. Estimates can be produced on a monthly basis at US level

Future work

- Improving breeding herd estimates
- Modeling spatial relationships among states

Selected References I

Bartlett, M. (1960). *Stochastic Population Models in Ecology and Epidemiology*. Methuen's monographs on applied probability and statistics. Methuen.

Beaumont, J. and Patak, Z. (2012). On the generalized bootstrap for sample surveys with special attention to poisson sampling. *International Statistical Review*, 80(1):127–148.

Box, G., Jenkins, G., Reinsel, G., and Ljung, G. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Bridges, T., Turner, L., Smith, E., Stahly, T., and Loewer, O. (1986). A mathematical procedure for estimating animal growth and body composition. *Transactions of the ASAE*, 29(5):1342–1347.

Busselberg, S. (2011). Bridging livestock survey results to published estimates through state–space models: A time series approach. *Proceedings of the Government Statistics Section, Joint Statistical Proceedings*.

Busselberg, S. (2013). The use of signal filtering for hog inventory estimation. *Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference*.

Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Eilers, P. and Marx, B. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, pages 89–102.

Fan, J., Wang, W., and Zhong, Y. (2019). Robust covariance estimation for approximate factor models. *Journal of econometrics*, 208(1):5–22.

Field, T., Taylor, R., and Field, T. (2003). Scientific farm animal production.

Flanders, F. and Gillespie, J. (2015). *Modern livestock & poultry production*. Cengage Learning.

Fletcher, R. (1987). *Practical methods of optimization*. Wiley-interscience publication. Wiley.

Selected References II

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, 20(1):5–10.

Horvitz, D. and Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685.

Hyndman, R., King, M., Pitrun, I., and Billah, B. (2005). Local linear forecasts using cubic smoothing splines. *Australian & New Zealand Journal of Statistics*, 47(1):87–99.

Hyndman, R. and Koehler, A. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.

Kedem, B. and Pan, L. (2015). *Time series prediction of hog inventory*. Unpublished internal document, United States Department of Agriculture NASS.

Khan, A. and Hildreth, W. (2003). *Case studies in public budgeting and financial management*. Marcel Dekker Nova York.

Lee, J., Embry, C., Hutter, J., Pollock, J., Rudd, R., Westrom, L., and Bull, A. (1996). Introduction to livestock and poultry production.

Little, R. and Rubin, D. (2014). *Statistical analysis with missing data*, volume 333. John Wiley & Sons.

Mashreghi, Z., Haziza, D., and Léger, C. (2016). A survey of bootstrap methods in finite population sampling. *Statistics Surveys*, 10:1–52.

Mikesell, R. and Baker, M. (2010). *Animal Science Biology and Technology*. Cengage Learning.

NASS (2005). *Estimation manual volume 4: livestock and dairy*. Unpublished internal document, United States Department of Agriculture NASS.

Park, H. and Oh, S. (2017). Seasonal variation in growth of berkshire pigs in alternative production systems. *Asian-Australasian Journal of Animal Sciences*, 30(5):749–754.

Selected References III

Park, H., Spann, K., Whitley, N., and Oh, S. (2017). Comparison of growth performance of berkshire purebreds and crossbreds sired by hereford and tamworth breeds raised in alternative production system. *Asian-Australasian Journal of Animal Sciences*, 30(9):1358–1362.

Pollard, J. (1966). On the use of the direct matrix product in analysing certain stochastic population models. *Biometrika*, 53(3-4):397–415.

Roberts, S. and Nowak, G. (2014). Stabilizing the lasso against cross-validation variability. *Computational Statistics & Data Analysis*, 70:198–211.

Rumelhart, D., Hinton, G., and Williams, R. (1985). Learning internal representations by error propagation. Technical report, California University San Diego, La Jolla Institute for Cognitive Science.

Shull, C. (2013). *Modeling growth of pigs reared to heavy weights*. PhD thesis, University of Illinois at Urbana-Champaign.

Tang, M., Cape, J., and Priebe, C. (2017). Asymptotically efficient estimators for stochastic blockmodels: the naive MLE, the rank-constrained MLE, and the spectral. *arXiv preprint arXiv:1710.10936*.

Thompson, P., Brewer, J., and Brewer, E. (1996). Swine care handbook. *National Pork Board*, page 34.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Wang, H., Li, G., and Tsai, C.-L. (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):63–78.

Thank you!

Questions?

Luca Sartore, PhD	lsartore@niss.org
Yijun Wei	ywei@niss.org
Emilola Abayomi, PhD	emilola.abayomi@usda.gov
Seth Riggins	seth.riggins@usda.gov
Gavin Corral, PhD	gavin.corral@usda.gov
Nell Sedransk, PhD	nsedransk@niss.org